Research Article

Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates.

Evgueni M. Andreev[1]

Vladimir M. Shkolnikov[2]

Alexander Z. Begun[3]

Abstract

A general algorithm for the decomposition of differences between two values of an aggregate demographic measure in respect to age and other dimensions is proposed. It assumes that the aggregate measure is computed from similar matrices of discrete demographic data for two populations under comparison. The algorithm estimates the effects of replacement for each elementary cell of one matrix by respective cell of another matrix. Application of the algorithm easily leads to the known formula for the age-decomposition of differences between two life expectancies. It also allows to develop new formulae for differences between healthy life expectancies. In the latter case, each age-component is split further into effects of mortality and effects of health. The application of the algorithm enables a numerical decomposition of the differences between total fertility rates and between parity progression ratios by age of the mother and parity. Empirical examples are based on mortality data from the USA, the UK, West Germany, and Poland and on fertility data from Russia.

[1]   Institute for Prognosis of the National Economy, Centre of Demography and Human Ecology, Nakhimovsky prosp., 47, 117418, Moscow; E-mail: evg_andreev@ns.cnt.ru

[2]   Max Planck Institute for Demographic Research, Doberaner Str. 114, D-18057 Rostock, Germany. E-mail: shkolnikov@demogr.mpg.de

[3]   Universität der Bundeswehr Hamburg, Holstenhofweg 85, 22043 Hamburg; E-mail: begun@unibw-hamburg.de

1. Introduction

In the 1980s, interest in decomposition appeared in relation to the particular problem of the decomposition of differences between two life expectancies. Nowadays, researchers are trying to develop general approaches which would solve the decomposition problem for wider classes of demographic measures (Das Gupta, 1994, Das Gupta, 1999, Vaupel and Canudas Romo, 2002, Shkolnikov, Andreev and Begun, 2001). The present study contributes to this line of research.

A variety of aggregate measures can be computed from demographic tables. Each of them aggregates a vector or a matrix of elementary rates of demographic events into one number. When analyzing changes in an aggregate demographic measure in time or its variations across countries, it is useful to be able to decompose observed changes or differences by age and other demographic dimensions such as birth order, cause of death, or population group. Decomposition aims at estimating contributions of differences between elementary rates of demographic events to the overall difference between two values of the aggregate measure.

This task is easier when analyzing for differences between two linear aggregates of elementary rates like two age-standardized rates or two total fertility rates. However, some of the aggregate measures are linked to elementary rates in a complex way. For example, life expectancy at birth is a functional of the vector of age-specific death rates, which has to be computed by complex acccumulation of these rates by means of the life table. The conventional *TFR* is simply the sum of age-specific fertility rates. However, *TFR* computed from age- and parity-specific fertility rates (denoted hereafter as *TFR_P*) is a complex measure, which has to be computed by means of the parity progression table.

A discrete method for decomposition of a difference between two life expectancies was independently developed in the 1980s by three different researchers from Russia, the USA, and France (Andreev, 1982, Arriaga, 1984, Pressat, 1985). The formulae for decomposition by Andreev and Pressat are exactly equivalent. Arriaga's formula is written in a slightly different form, but it is essentially equivalent to the formulae by Andreev and Pressat (Shkolnikov et al., 2001). A continuous version of the method for decomposition of differences between life expectancies by age was developed by Pollard (1982).

It appears that existing formulae for age-decomposition of the difference between life expectancies are just particular forms of a general algorithm, which includes the *stepwise replacement* of elements from one vector of age-specific mortality rates by respective elements of another vector (Shkolnikov, Andreev and Begun, 2001). In a more general case, elements of one multidimensional matrix should be replaced by

respective elements of another matrix. This approach has its roots in the general idea of standardization (Kitagawa, 1964).

The first section of the present paper is devoted to the decomposition of differences between life expectancies and between healthy life expectancies. First, we show how the conventional formula for decomposition of differences between life expectancies derives from the general replacement algorithm. Second, the same approach is applied to the age-decomposition of differences between health expectancies. These measures combine data on mortality with data on health. Each measure is calculated from two independent vectors, that is, of age-specific mortality rates and of the age-specific prevalences of "good" health. Correspondingly, each elementary age-component of the difference between two health expectancies is to be split further into the effects of mortality and of health.

The second section deals with two aggregate measures of fertility based on the parity-progression table, namely parity-progression ratios and the total fertility rate. Both indicators are based on the matrix of elementary fertility rates by age of the mother and parity. We show that the results of an exact decomposition according to the algorithm of stepwise replacement differ from those returned by simpler methods.

In sections 1 we solve the decomposition problems by developing formulae for the components. In section 2 we describe a procedure for their numerical estimation.

2.   Decomposition of differences between two life expectancies and between two healthy life expectancies

Life expectancies

Consider two life expectancies at birth $e_0^1 = e_0(M^1)$ and $e_0^2 = e_0(M^2)$ computed in a conventional way from two vectors of age-specific mortality rates $M^1 = \left\| m_x^1 \right\|$ and $M^2 = \left\| m_x^2 \right\|$, $x=0, 1, 2, \ldots \omega$. Transition from life expectancy e_0^1 to life expectancy e_0^2 corresponds to a transformation of vector M^1 into vector M^2. Such a transformation can be completed by a *stepwise replacement* in an age-by-age mode of the elements of the first vector by respective elements of the second vector.

Let $M^{[x]}$ be the vector, consisting of elementary mortality rates ${}_1 m_y^2$ at ages $y < x$ and ${}_1 m_y^1$ at ages $y \geq x$ (Note 1). The difference $\delta_{0|x}^{2-1} = e_0(M^{[x]}) - e_0^1$ is a contribution of ages from $y < x$ to the overall difference $e_0^2 - e_0^1$. Using definitions of the standard life tables functions, it can be presented as

$$\delta_{0|x}^{2-1} = e_0(M^{[x]}) - e_0^1 = (L_{0|x}^2 - L_{0|x}^1) + (l_x^2 - l_x^1) \cdot e_x^1, \tag{1}$$

where $L_{0|x} = \int_0^x l(t)dt$. The first additive term in (1) is the length-of-life effect of replacement 1→2 at ages under x, the second additive term is the effect of 1→2 replacement at ages under x on life expectancy after age x.

The contribution of elementary age interval $[x, x+1)$ can be expressed as

$$\delta_x^{2-1} = \delta_{0|x+1}^{2-1} - \delta_{0|x}^{2-1} = e_0(M^{[x+1]}) - e_0(M^{[x]}) \tag{2}$$

Formula (2) is a form of the general algorithm of stepwise replacement for a one-dimensional decomposition of a difference between two aggregate measures by age.

Substitution of (2) in (1) yields

$$\delta_x^{2-1} = l_x^2(e_x^2 - e_x^1) - l_{x+1}^2(e_{x+1}^2 - e_{x+1}^1).$$

Thus, the overall difference between two life expectancies is

$$e_0^2 - e_0^1 = \sum_{x=0}^{\omega} \delta_x^{2-1}, \tag{3a}$$

where $l_{\omega+1}$ and $e_{\omega+1}$ are assumed to be 0.

In a similar way, one can decompose the difference $e_0^1 - e_0^2$ by making 2→1 replacements instead of 1→2 replacements:

$$e_0^1 - e_0^2 = \sum_{x=0}^{\omega} \delta_x^{1-2} = \sum_{x=0}^{\omega} [l_x^1(e_x^1 - e_x^2) - l_{x+1}^1(e_{x+1}^1 - e_{x+1}^2)]. \tag{3b}$$

Components δ_x^{2-1} are somewhat different from components $-\delta_x^{1-2}$. That is to say that the decomposition of differences between life expectancies depends on the permutation of vectors under

$$e_0^2 - e_0^1 = \sum_{x=0}^{\omega} \delta_x = \frac{1}{2} \cdot \sum_{x=0}^{\omega} \{[l_x^2(e_x^2 - e_x^1) - l_{x+1}^2(e_{x+1}^2 - e_{x+1}^1)] - [l_x^1(e_x^1 - e_x^2) - l_{x+1}^1(e_{x+1}^1 - e_{x+1}^2)]\} \quad (3)$$

comparison. Both E.Andreev (1982) and R.Pressat (1985) noticed this important feature. They suggested averaging as a way to obtain symmetrical components i.e.

$$\delta_x = \frac{1}{2} \cdot (\delta_x^{2-1} - \delta_x^{1-2}) \cdot$$

The latter expression exactly replicates the formulae by E. Andreev (1982) and R. Pressat (1985).

Numerous empirical examples of decompositions of differences between life expectancies can be found elsewhere (Shkolnikov et al., 2001).

Replacement running from young to old ages looks natural and meaningful. Nevertheless, it is not obvious why the replacement algorithm goes this way. Generally speaking, it could be organized differently. For example, it could run from old to young ages (Pollard, 1988) or in a random manner.

The most general procedure for the replacement of one element should include $2^{\omega-1}$ replacements. For every age x the replacement should be performed with all possible combinations of M_y^1 and M_z^2, at ages y and z other than age x. The final contribution of age x should be calculated as the average of $2^{\omega-1}$ components.

This procedure is extremely laborious given high numbers of age groups (about 20 for abridged life tables and about 100 for complete ones). Several numerical experiments (not shown here) suggest that results of "complete" decomposition of differences between life expectancies are close to the results returned by formula (3). So, there are reasons to avoid laborious calculations if we agree to follow the existing tradition of making replacements in ascending order of ages. It guarantees consistency regarding the existing formulae for the decomposition of differences between life expectancies by age and, as we show below, allows to develop similar decomposition formulae for the age-decomposition for other aggregate measures such as Gini coefficients or healthy life expectancies.

Health expectancies

Health expectancy indicators can be built in several ways (Cambois, Robine and Brouard, 1999, Robine, Romieu and Cambois, 1999). So far, the "observed prevalence life table method" by D.Sullivan (1964) is the most widely used method. According to this method, health expectancy is defined as

$$h_0 = \sum_{x=0}^{\omega} {}_1 L_x \cdot \pi_x, \tag{4}$$

where π_x is the share of person-years lived in "good" health within the elementary age interval $[x, x+1)$. Usually the health-weights π_x are obtained from nationally representative surveys including questions on self-perceived health, self-perceived disability, ill-health or physical performance scales (Wilkins and Adams, 1983, Robine, Romieu and Cambois, 1999, Doblhammer and Kityr, 2001, Crimmins and Saito, 2001).

According to (4), two vectors are needed for calculating the health expectancy. These are the vector of age-specific mortality rates M and the vector of age-specific health-weights Π. Correspondingly, decomposition of the difference between two health expectancies should include additional splitting of each age-component into effects of mortality and health.

According to the algorithm of stepwise replacement, the component of the overall difference in h_0 due to the difference between mortality rates at age x is

$$\lambda_x^{2-1} = \frac{1}{2} \{ [(h_0(M^{[x+1]}, \Pi^{[x]}) - h_0(M^{[x]}, \Pi^{[x]})] + [(h_0(M^{[x+1]}, \Pi^{[x+1]}) - h_0(M^{[x]}, \Pi^{[x+1]})] \} \tag{5}$$

The component of the overall difference in h_0 due to the difference in health-weights at age x is

$$\gamma_x^{2-1} = \frac{1}{2} \{ [h_0(M^{[x]}, \Pi^{[x+1]}) - h_0(M^{[x]}, \Pi^{[x]})] + [h_0(M^{[x+1]}, \Pi^{[x+1]}) - h_0(M^{[x+1]}, \Pi^{[x]})] \} \tag{6}$$

Expressions (5) and (6) allow to estimate components λ_x^{2-1} and γ_x^{2-1} numerically. It is also possible to develop formulae for them. Let ${}_1 P_x^i = \frac{{}_1 L_x^i}{l_x^i}$ (i=1,2). Then formula (4) can be re-written as

$$h_0 = \sum_{y=0}^{x-1}(L_y\pi_y) + {}_1P_x\ \pi_x l_x + h_{x+1}l_{x+1} \tag{7}$$

The first term corresponds to ages younger than *x*, the second term corresponds to age [*x, x*+1) and the third term corresponds to ages older than *x*. The second term is equal to the number of person-years in good health lived at age *x*. It also consists of three components: l_x depends on mortality at ages younger than *x*, ${}_1P_x$ depends on mortality at age *x* and π_x depends on the prevalence of good health at age *x*.

1→2 replacement of mortality rate at age *x* produces changes in the second and the third terms of (7). 1→2 replacement of health-weight at age *x* produces change in the second term of (7).

Insertion of (7) in (5) and (6) leads to

$$\lambda_x^{2-1} = \frac{1}{2}\{[(l_{x\ 1}^2 P_x^2 \pi_x^1 + h_{x+1}^1 l_x^2(1-{}_1q_x^2)) - (l_{x\ 1}^2 P_x^1 \pi_x^1 + h_{x+1}^1 l_x^2(1-{}_1q_x^1)] +$$

$$[(l_{x\ 1}^2 P_x^2 \pi_x^2 + h_{x+1}^1 l_x^2(1-{}_1q_x^2)) - (l_{x\ 1}^2 P_x^1 \pi_x^2 + h_{x+1}^1 l_x^2(1-{}_1q_x^1)]\}$$

or

$$\lambda_x^{2-1} = \frac{1}{2}l_x^2\left({}_1P_x^2 - {}_1P_x^1\right)\left({}_1\pi_x^1 + {}_1\pi_x^2\right) + h_{x+1}^1 l_x^2\left({}_1q_x^1 - {}_1q_x^2\right).$$

In a similar way the health-related component at age *x* is

$$\gamma_x^{2-1} = \frac{1}{2}[l_{x\ 1}^2 P_x^1 \pi_x^2 - l_{x\ 1}^2 P_x^1 \pi_x^1) + (l_{x\ 1}^2 P_x^2 \pi_x^2 - l_{x\ 1}^2 P_x^2 \pi_x^1)]$$

or

$$\gamma_x^{2-1} = \frac{1}{2}l_x^2\left({}_1P_x^1 + {}_1P_x^2\right)(\pi_x^2 - \pi_x^1)$$

The components $\lambda_x^{2-1}, \gamma_x^{2-1}$, produced by 1→2 replacement, should be averaged with $\lambda_x^{1-2}, \gamma_x^{1-2}$ produced by 2→1 replacement. The final formulae are

$$\lambda_x = \frac{1}{4}(l_x^1 + l_x^2)(_1P_x^2 -_1P_x^1)(\pi_x^1 + \pi_x^2) + \frac{1}{2}(h_{x+1}^1 l_x^2 + h_{x+1}^2 l_x^1)(_1q_x^1 -_1q_x^2) \quad (8)$$

$$\gamma_x = \frac{1}{4}(l_x^1 + l_x^2)(_1P_x^1 +_1P_x^2)(\pi_x^2 - \pi_x^1) \quad (9)$$

Age-components λ_x of the difference between health expectancies derived from (10) are equivalent to age-components of the difference between life expectancies δ_x from (3) if $\pi_x^1 = \pi_x^2 = 1$ for all ages x=0,1,2, ..., ω .

The example in Figure 1 shows age-mortality-health-specific components of the difference between female health expectancies at age 20 in West Germany and Poland. Health-weights are calculated from the data on self-perceived health, extracted from the second and third wave of the World Value Surveys (Inglehart et al., 2000). For each five-year age group weights π_x are the sums of the original proportions of women with "fair", "good" and "very good" self-perceived health. For both countries the mortality rates for the year 1995 are used. From these data health expectancies at age 20 are computed for West Germany and Poland and their difference is decomposed by age according to (8) and (9).

Figure 1 suggests that contributions due to differences in self-perceived health are much greater than those due to differences in mortality. Indeed, 7.2 years of the overall difference of 8.9 years are attributable to differences in health. Although the maximum age-specific contributions are produced for ages from 65 to 75, they are very significant as early as at the age of 50.

It is worth understanding that the mortality age-components of the difference between health expectancies λ_x are different from age-components δ_x of the difference between respective life expectancies if for some ages x $\pi_x \neq 1$. Indeed, in the example given above, female life expectancies at age 20 in West Germany and Poland were 61 and 58 years, respectively. The difference of 3 years is significantly higher than the total effect of mortality in the West Germany-Poland difference between health expectancies, which is equal to 1.7 years.

Figure 1: *Decomposition of the difference between female health expectancies at age 20 between West Germany and Poland for the mid-1990s.*
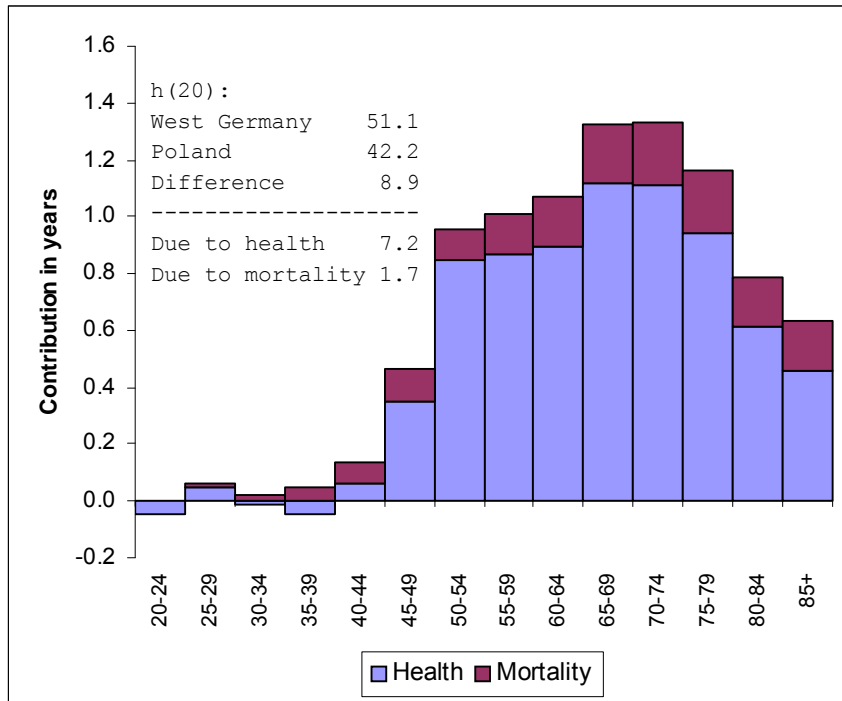
3. Decomposition of differences between two parity progression ratios and two TFRs computed from age-parity-specific fertility rates

The general replacement algorithm can be applied to differences between aggregate fertility measures based on parity-progression tables. In this section we apply the general algorithm of stepwise replacement to estimate the age- and parity-specific components of differences between two parity-progression ratios and of differences between two total fertility rates.

A conventional *TFR* based on age-specific fertility rates is adjusted for population age structure but ignores differences between sequences of births. The *TFR_P* based on age-parity-specific fertility rates is adjusted both for age and parity. This type of period fertility measure is valuable for analytical purposes (Kohler and Ortega, 2002, Barkalov, 1999, Darsky and Scherbov, 1990).

Let us consider an example of changes in Russian fertility in the first half of the 1990s. The *TFR* in this country dropped from 2 in 1989 to 1.4 in 1993-94 (Table 1). Table 1 shows also that in these years the difference between the *TFR* and the *TFR_P* in Russia increased.

Table 1:   *Total fertility rates computed without and with data on parity in Russia in 1989-1994.*

Year	TFR	TFR_P
1989	2.01	1.97
1990	1.89	1.87
1991	1.73	1.75
1992	1.55	1.62
1993	1.39	1.45
1994	1.40	1.46

Sources: Andreev and Barkalov, 1999
                    Goskomstat, 2001

The parity-progression table is a type of a multi-status demographic table based on the matrix $F = \left\| f_{x,par} \right\|$, in which each elementary fertility rate $f_{x,par}$ is a ratio of the number of *par*-order births to the mid-year population of women aged *x* with *par*-1 children. In matrix *F,* ages are presented as rows and birth orders are presented as columns. Age runs across the interval of reproductive ages $[\alpha, \beta]$, while parity varies from 0 to the highest birth order *p*.

The parity-progression table can be computed from matrix *F* (Whelpton, 1946, Chiang and Van Den Berg, 1982, Lutz, 1989, Andreev and Barkalov, 1999, Kohler and Ortega, 2002). The computational procedure, used in the present study, is given in the Appendix.

The total parity-specific birth numbers are defined as $B_0 = l_{\alpha,0} = 1000$ for *par*=0 and $B_{par} = \sum_{x=\alpha}^{\beta} b_{x,par}$ for *par*>0, where $b_{x,par}$ are age-parity-specific birth numbers from the parity-progression table.

Two aggregate fertility measures, which will be used for our next decomposition exercises, are the

parity-progression ratio

$$a_{par} = B_{par+1} / B_{par} \tag{10}$$

and the total fertility rate

$$TFR\_P = B_0 \cdot \sum_{par=1}^{p} B_{par} = B_0 \cdot (a_0 + a_0 a_1 + \ldots + a_0 a_1 a_2 \ldots a_{p-1}) \tag{11}$$

Formulae (10), (11) and Appendix define a_{par} and *TFR_P* as functionals of the matrix of elementary fertility rates *F*.

At first glance, (10) and (11) suggests that the decompositions of difference between values of *TFR_P* in respect to age *x* or parity *par* can be accomplished in a simple way from $b_{x,\bullet}$ values and a_{par} values, respectively. N.Barkalov (1999) developed a formula for the components of differences between two *TFR_P* values. This formula corresponds to the stepwise replacement of a_{par}^1 by a_{par}^2 in (11) running from lower to higher parities, although the way of its development by the author was different. It appears, however, that a_{par} depends not only on $f_{x,par+1}$, but also on fertility rates of lower parities $f_{x,k}$ (*k*<*par*+1). We will show later that numerical results of this method of decomposition are somewhat different from the exact results of the general replacement algorithm.

The general replacement algorithm for decomposition of differences between a_{par} values or *TFR_P* values includes replacements of all elementary fertility rates in matrix F^1 by respective elements of matrix F^2 and vice versa. The external cycle of replacement runs across ages, while the internal cycle runs across parities. As in the previous section, the replacement runs in ascending order by age. It means that the

replacement of the elements of row x $f^1_{x,par}$ by elements $f^2_{x,par}$ begins only after all elements $f^1_{y,par}$ are already replaced by the elements $f^2_{y,par}$ for rows $y<x$. Certainly, the replacement could be organized differently with parity being the "first" dimension and age being the "second" one (see section 3).

The effect of replacement of a given element $f^1_{x,par}$ by $f^2_{x,par}$ on a_{par} and on *TFR_P* should be calculated 2^{par-1} times according to the number of different combinations of $f^1_{x,k}$ with $f^2_{x,l}$ in remaining cells of the same row x ($k < p, l < p$ and $k \neq l \neq par$). Finally, the component produced by age x and parity *par* in replacement $1\rightarrow2$ is the average of all 2^{par-1} effects.

As in section 1, in order to obtain symmetrical components the whole set of replacements should be completed twice in two directions ($1\rightarrow2$ and $2\rightarrow1$).

Let us consider another example. Data from the Russian census of 1989 and the micro-census of 1994 allow us to estimate the composition of female population by age and parity. These data combined with annual statistics on births by parity and age of the mother allow us to calculate fertility rates by parity (1, 2, ..., 5+) and age of the mother (15, 16, 17, ..., 54). Thus, we have to operate with two matrices (40 x 5) of elementary fertility rates F^1 and F^2. Values of the parity progression ratios, computed from the matrices of 1989 and 1994, show a structure of the rapid drop in Russian fertility between 1989 and 1994 (Table 2).

Table 2: *Changes in parity progression ratios and in total fertility rate between 1989 and 1994.*

Year	a^{1989}_{par-1} and a^{1994}_{par-1}					TFR_P^{1989} and TFR_P^{1994}
	1	**2**	**3**	**4**	**5+**	
1989	0.948	0.767	0.293	0.269	0.519	1974
1994	0.895	0.530	0.161	0.182	0.317	1464
1994-1989	-0.053	-0.237	-0.132	-0.087	-0.202	-510

Replacement of a single element of matrix F^1 by the respective element of matrix F^2 (or vice versa) includes 16 (2^4) calculations of changes in a_{par} (*par*=1,2,...,5+) and in

TFR_P. For example, replacement of the upper-left element $f_{15,1}^{1}$, corresponding to age 15 and parity 1, by the element $f_{15,1}^{2}$ includes replacements of this element in all possible combinations with rates $f_{15,k}^{1}$ and $f_{15,l}^{2}$ in the remaining 4 cells of the same row with $k \neq l$ and $k,l \in [2,3,4,5+]$. These calculations return 16 magnitudes of changes in the output measures (a_{par} and *TFR_P*) produced by the replacements. Finally, the averages of these 16 values should be computed for each a_{par} and *TFR_P*.

Tables 3a, 3b, and 3c show age- and parity-specific contributions to the overall changes in a_1, a_2 and a_3 in Russia between 1989 and 1994. In general, decreases in progressions to a given parity *par* are mostly due to decreases in fertility rates for the same parity *par*. However, changing fertility rates at lower parities (*par*-1 or even *par*-2) also produce some effects on progression to parity *par*. These effects are more significant for higher parities (e.g. progressions to the third and the fourth births).

Figure 2 shows the structure by age and parity of the decrease in *TFR_P* in Russia between 1989 and 1994. The greatest contributions are produced by decreases in second births by mothers at ages from 23 to 29.

Table 3a: *Components of decline in progression to second births by age and parity, Russia, 1989-1994.*

Age group	a_1 progression to second births		
	1989:	0.77	
	1994:	0.53	
	Difference:	-0.24	
	All parities	**1**	**2**
15-19	0.00	0.00	0.00
20-24	-0.06	-0.01	-0.05
25-29	-0.08	0.00	-0.08
30-34	-0.07	0.00	-0.07
35-39	-0.02	0.00	-0.03
40-54	0.00	0.00	0.00
All ages	-0.24	-0.01	-0.23

Table 3b: *Components of decline in progression to third births by age and parity, Russia, 1989-1994.*

Age group	a_2 progression to third births			
	1989:	0.29		
	1994:	0.16		
	Difference:	-0.13		

	All parities	1	2	3
15-19	0.00	0.00	0.00	0.00
20-24	-0.04	0.00	-0.03	-0.01
25-29	-0.04	0.00	-0.01	-0.03
30-34	-0.03	0.00	0.01	-0.04
35-39	-0.02	0.00	0.01	-0.03
40-54	0.00	0.00	0.00	0.00
All ages	-0.13	0.00	-0.02	-0.11

Table 3c: *Components of decline in progression to fourth births by age and parity, Russia, 1989-1994.*

Age group	a_3 progression to fourth births				
	1989:	0.27			
	1994:	0.18			
	Difference:	-0.09			

|	All parities	1	2	3	4
15-19	0.00	0.00	0.00	0.00	0.00
20-24	-0.03	0.00	-0.02	-0.01	0.00
25-29	-0.02	0.00	0.00	-0.01	-0.02
30-34	-0.01	0.00	0.01	0.01	-0.03
35-39	-0.01	0.00	0.00	0.02	-0.03
40-44	-0.01	0.00	0.00	0.00	-0.01
45-54	0.00	0.00	0.00	0.00	0.00
All ages	-0.09	0.00	-0.01	0.02	-0.10

As we mentioned before, formulae (10) and (11) give the impression that decompositions can be made in simpler ways. Indeed, according to (11), *TFR_P* is the total of all age- and parity-specific birth numbers $b_{x,par}$. Thus, it seems that age-components of the change in *TFR_P* could be calculated simply as

$$b_{x,\bullet}^2 - b_{x,\bullet}^1 = \sum_{par=1}^{p} b_{x,par}^2 - \sum_{par=1}^{p} b_{x,par}^1$$. This approach ignores the dependence of $b_{x,par}^i$

values from fertility rates at parities lower than *par*. The result of such a calculation is substantially biased in comparison to the exact age-components returned by the general replacement algorithm (Figure 3).

Figure 2: *Decomposition of decrease in the total fertility rate in Russia in 1989-1994 by age and parity.*
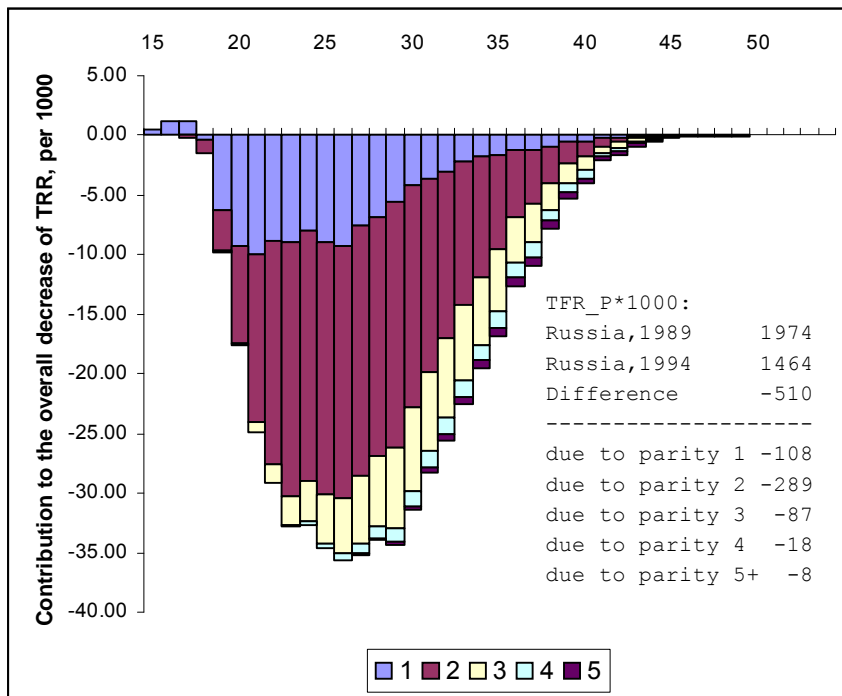
Figure 3: *Age-components of decrease in* TFR_P *between 1989 and 1994 in Russia. A comparison of the exact decomposition by the replacement algorithm with a simplified decomposition according to differences in birth numbers from parity progression table.*
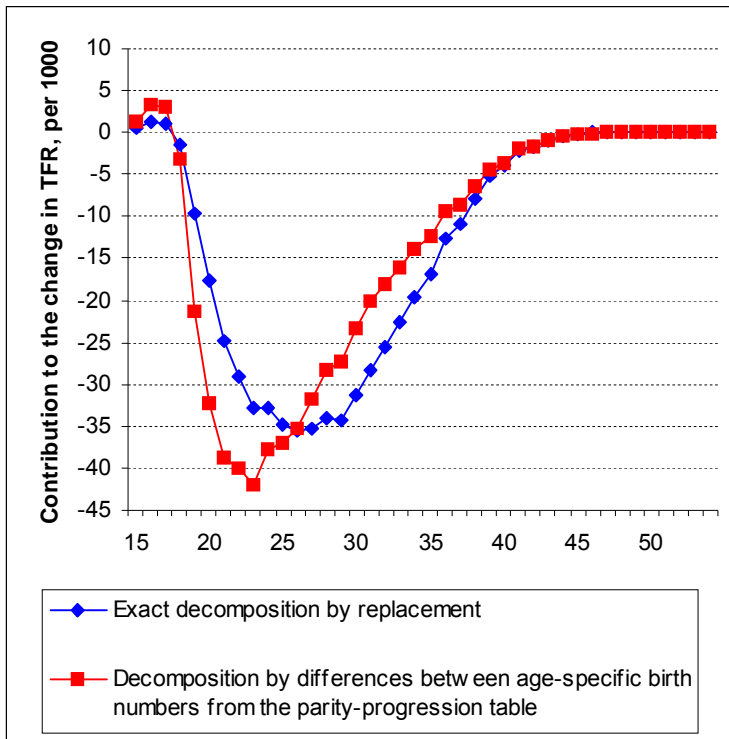
A simplified decomposition of a difference between *TFR_P* values in respect to parity can be accomplished in two different ways. The simplest is to imagine that the *par*-component of difference between two values of *TFR_P* is $B_{par}^2 - B_{par}^1$. This approach ignores the dependence of higher-order birth numbers on lower-order births and results in significantly biased parity-components (Table 4). The method by Barkalov, mentioned above, returns the *par*-components, which are rather close but not equivalent to exact components returned by the general replacement algorithm. (Table 4).

Table 4:   *Components of decline in* TFR_P *by parity computed by three methods, Russia, 1989-1994 (per 1000)*

Method	Parity					Total change
	1	2	3	4	5+	
$B_{par}^2 - B_{par}^1$	-53	-252	-136	-43	-25	510
Method by Barkalov: stepwise replacement of a_{par} in (13)	-97	-286	-107	-15	-5	510
General replacement algorithm: stepwise replacement of $f_{x,par}$	-108	-289	-87	-18	-8	510

4. Final considerations

The algorithm of stepwise replacement is a universal tool for the decomposition of differences between aggregate measures computed from demographic tables. The number of examples in the present study could be increased. In particular, in an earlier study (Shkolnikov, Andreev and Begun, 2001) we developed a new formulae for the age-decomposition of differences between two Gini coefficients (measures of variability in age at death). In this study we showed also how to decompose the age-components further in respect to age and population composition by social group. Such decomposition can be accomplished by means of the same general algorithm including replacement of age-group-specific death rates and of age-specific population weights of groups.

We would like to add two additional comments regarding two peculiar aspects of the decomposition which were not addressed in sections 1 and 2, but should be kept in mind when making decompositions and interpreting their results.

Path dependence. In the present paper, we always assume that population "jumps" from state 1 to state 2 with no intermediate states in between. However, results of the decomposition could depend on a particular pathway of transition from state 1 to state 2. For example, the age-parity components of change in the Russian *TFR_P* from 1989 to 1994 are somewhat different from the sums of age-parity components of subsequent transitions 1989→1990, 1990→1991, 1991→1992, 1992→1993, and 1993→1994.

The best way to address this problem is to make all annual transitions and then to sum them up. However, this solution is not really popular among researchers because

differences between a direct transition (as 1989→1994) and the sequence of annual transitions are usually small. However, there is no guarantee that such differences would be small in all cases.

Age as a special dimension. In the present paper we treated age as a special dimension in two respects. First, replacement ran from younger to older ages in order to keep a consistency with earlier formulae by Andreev (1982), Arriaga (1984) and Pressat (1985). This approach gives also an opportunity for developing new formulae for other aggregate measures. Second, we were interested in splitting further each age-component according to additional dimensions (effects of mortality and health or effects of parities within each age group). This means that age always played the role of the first dimension. The formulae developed in sections 1 and 2 *correspond to this particular approach*.

Dependence on the sequence of replacement. The order of dimensions in the course of replacement could also matter for results. For example, there are two ways to replace the age-group-specific mortality rates and age-specific population-weights of groups when decomposing a difference between two life expectancies from data on mortality by age and population group. One can make a replacement of age-specific mortality rates within each population group or to replace group-specific mortality rates within one age group. Generally speaking, all replacement schemes are equally acceptable. Ideally, the final components should be based on the averaging of effects produced by all possible sequences of dimensions (Das Gupta, 1994, 1999). This general principle works well for linear aggregates, but could lead to long computational times for more complex measures, especially if the number of dimensions and the number of categories within each dimension is large.

Notes

1. For the sake of simplicity, notation in all formulae of this paper are given for complete demographic tables (tables with single-year age groups). All of them can be easily re-written for abridged demographic tables.

Corrections:
On February 27, 2018, two equations on page 505 were updated at the authors' request.

References

Andreev, E.M. (1982). Metod komponent v analize prodoljitelnosty zjizni. [The method of components in the analysis of length of life]. *Vestnik Statistiki*, 9, 42-47.

Andreev, E.M. and Barkalov, N. (1999). Birth tables by sequence of birth. [Tablitsy rozhdaemosti po ocherednosti rozhdenii.] *Voprosy Statistiki*, 5, 64-67.

Arriaga, E. (1984). Measuring and explaining the change in life expectancies. *Demography* 21(1), 83-96.

Barkalov, N.B. (1999). The fertility decline in Russia, 1989-1996: a view with period parity-progression ratios. *Genus*, LV(3-4), 11-60.

Cambois, E., Robine, J.-M., Brouard, N. (1999). Life expectancies applied to specific statuses. A history of the indicators and the methods of calculation. *Population: An English Selection*, 11, 7-34.

Chiang, C.L., Van Den Berg, B.J. (1982). A fertility table for the analysis of human reproduction. *Mathematical Biosciences*, 62, 237-251.

Crimmins, E.M. and Saito, Y. (2001). Trends in healthy life expectancy in the United States, 1970-1990: gender, racial, and educational differences. *Social Science and Medicine*, 52, 1629-1641.

Darsky, L.E. and Scherbov, S. (1990). Parity-progression fertility tables for the nationalities of the USSR. *IIASA Working Paper*, WP-90-53, September.

Das Gupta, P. (1994). Standardisation and decomposing of rates from cross-classified data. *Genus*, L(3-4), 171-196.

Das Gupta, P. (1999). Decomposing the difference between rates when the rate is a function of factors that are not cross-classified. *Genus*, LV(1-2), 9-26.

Doblhammer, G. and Kytir, J. (2001). Compression or expansion of morbidity? Trends in healthy-life expectancy in the elderly Austrian population between 1978 and 1998. *Social Science and Medicine*, 52, 385-391.

Goskomstat (2001). *The Demographic Yearbook of Russia*. Goskomstat of the Russian Federation, Moscow.

Hanada, K. (1983). A formula of Gini's concentration ratio and its application to life tables. *Journal of Japan Statistical Society*, 13(2), 95-98.

Inglehart, R., et al. (2000). *World Value Surveys and European Value Surveys, 1981-1984, 1990-1993, and 1995-1997*. [Computer file]. ICPSR version. Ann Arbor, MI, Inter-University Consortium for Political and Social Research.

Kitagawa, E. (1964). Standardized comparisons in population research. *Demography*, 1, 296-315.

Kohler H.-P. and Ortega J.A. (2002). Tempo-adjusted period parity-progression measures: assessing the implications of delayed childbearing for cohort fertility in Sweden, the Netherlands and Spain. *Demographic Research*, 6, article 7, March.

Kohler, H.-P., Billari, F. and Ortega J.A. (2002). *Lowest-Low Fertility*. Book manuscript.

Lutz, W. (1989). *Distributional Aspects of Human Fertility: a Global Comparative Study*. New York: Academic Press.

Pollard, J.H. (1982). The expectation of life and its relationship to mortality. *Journal of the Institute of Actuaries*, 109, Part II, No 442, 225-240.

Pollard, J.H. (1988). On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography,* 25, 265-276.

Pressat, R. (1985). Contribution des écarts de mortalité par âge à la différance des vies moyennes. *Population*, 4-5, 766-770.

Robine J.-M., Romieu, I., Cambois, E. (1999). Health expectancy indicators. *Bulletin of the World Health Organisation*, 77(2), 181-185.

Shkolnikov V.M., Andreev E.M., Begun A.Z. (2001). Gini coefficient as a life table function: computation from discrete data, decomposition of differences and empirical examples. *MPIDR Working Paper*, WP-2001-017. Available at www.demogr.mpg.de

Shkolnikov, V., Valkonen, T., Begun, A., Andreev, E. (2001). Measuring inter-group inequalities in length of life. *Genus*, LVII(3-4), 33-62.

Sullivan, D.F. (1964). A single index of mortality and morbidity. *HSMHA health report*, 86, 347-354.

Vaupel, J.W. and Canudas Romo, V. (2002) Decomposing demographic change into direct vs. compositional components. Demographic Research, 7-1, www.demographic-research.org.

Whelpton, P.K. (1946) Reproduction rates adjusted for age, parity, fecundity and marriage. *Journal of the American Statistical Association*, 41, 501-516.

Wilkins, R. Adams, O.B. (1983). Health expectancy in Canada, late 1970s: demographic, regional and social dimensions. *American Journal of Public Health*, 73(9), 1073-1080.

Appendix

We used the following scheme for computation of the period parity-progression tables:

$$l_{\alpha,0} = 1000 \text{ and } l_{\alpha,par} = 0 \text{ for } par>0$$

$$\varphi_{x,par} = 2f_{x,par} / (2 + f_{x,par})$$

$$b_{x,1} = l_{x,0} \cdot \varphi_{x,0} \text{ and } b_{x,par} = (l_{x,par} + b_{x,par-1} / 2) \cdot \varphi_{x,par} \text{ for } par>0$$

$$l_{x+1,0} = l_{x,0} - b_{x,1} \text{ and } l_{x+1,par} = l_{x,par} - b_{x,par} + b_{x+1,par} \text{ for } par>0$$